Exhibit D

Part 5

CMU's Incorrect Construction Covers Euclidean Metric

- "Euclidean branch metric" has noise samples that *vary together* ["identically" with "variance σ^2"]

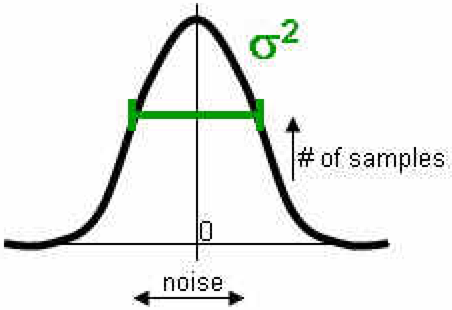
> Euclidian branch metric. In the simplest case, the noise samples are realizations of independent identically distributed Gaussian random variables with zero mean and variance σ^2. This is a white Gaussian noise assumption. This

'839 Patent 5:59-62

$$M_i = N_i^2 = (r_i - m_i)^2 \quad (8)$$

'839 Patent 6:12-13

Gaussian noise



- CMU argued to the Patent Office:

> $$M_i = [r_i(0) - y_i(0)]^2 + [r_i(1) - y_i(1)]^2$$
>
> Such a branch metric is not correlation sensitive, as claimed in independent claims 11, 16,

6/12/00 Amdt. at 9, '839 Patent File History (Marvell Exh. 22)

CMU's "Disclosed Embodiment" Argument Fails

- CMU concedes: "Figure 3A calculates a correlation … in one disclosed embodiment"

FIG. 3A illustrates a block diagram of a branch metric computation circuit **48** that computes the metric M_i for a branch of a trellis, as in Equation (13). Each branch of the

CMU Reply at 4

'839 Patent 7:14-18

- ► Figure 3A computes Equation (13)
- ► Using Noise Covariance Matrix $\hat{C}$
- ► The $\hat{C}(\hat{a})$ estimate calculates correlation: $E[\hat{C}(\hat{a})] = E[N_i N_i^T]$
- ► **Marvell covers this embodiment**

need for further mean corrections. The focus is shifted to tracking the noise covariance matrices needed in the computation of the branch metrics (13).

Assume that the sequence of samples $r_i, r_{i+1}, \ldots, r_{i+L}$ is observed. Based on these and all other neighboring samples, after an appropriate delay of the Viterbi trellis, a decision is made that the most likely estimate for the sequence of symbols $a_{i-K_l}, \ldots, a_{i+L+K_t}$ is $\hat{a}_{i-K_l}, \ldots, \hat{a}_{i+L+K_t}$. Here L is the noise correlation length and $K=K_l+K_t+1$ is the ISI length. Let the current estimate for the $(L+1)\times(L+1)$ covariance matrix corresponding to the sequence of symbols $\hat{a}_{i-K_l}, \ldots, \hat{a}_{i+L+K_t}$ be $\hat{C}(\hat{a}_{i-K_l}, \ldots, \hat{a}_{i+L+K_t})$.
This symbol is abbreviated with the shorter notation, $\hat{C}(\hat{a})$. If the estimate is unbiased, the expected value of the estimate is:

$$E\hat{C}(\hat{a})=E[\underline{N}_i\underline{N}_i^T] \tag{21}$$

where $\underline{N}_i$ is the vector of differences between the observed samples and their expected values, as defined in (12).

'839 Patent 9:21-37

CMU's "Disclosed Embodiment" Argument Fails

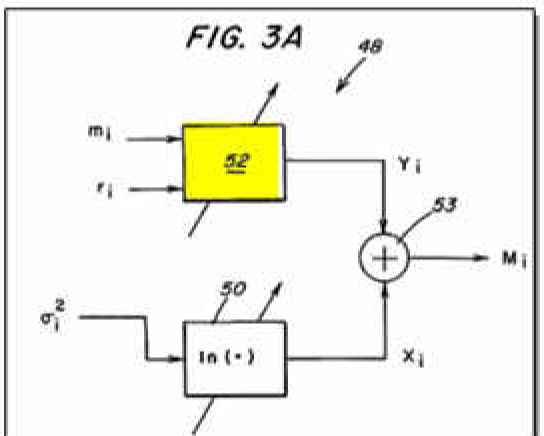
- McLaughlin: "an embodiment of which is shown in Figure 3B of the CMU Patents … without using or computing the expected value of the product of the signal samples."

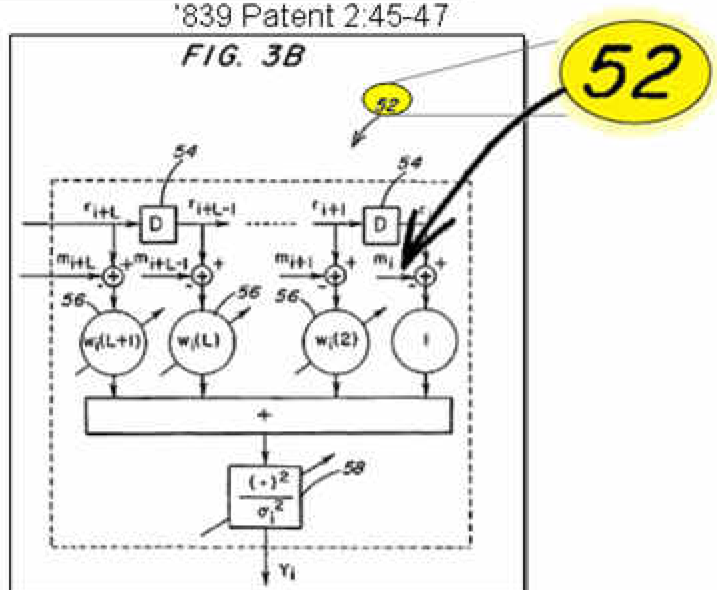
McLaughlin Decl. at ¶ 13

1. Fig. 3B only shows one implementation of Circuit 52 in Fig. 3A

> FIG. 3B is an illustration of an implementation of a portion of the branch metric computation module of FIG. 3A;

'839 Patent 2:45-47





Covariance

Claim Term	CMU's Construction	Marvell's Construction
covariance '839 Patent Claims 11, 16, 19, 23 '180 Patent Claim 6	none; see "noise covariance matrices" CMU Brf. at 27	the expected (mean) value of the product of (r_i-m_i) and (r_j-m_j), where r_i and r_j are observed signal samples (at time I and time j, respectively) and m_i and m_j are the expected (mean) values of the samples (at time i and time j, respectively) (i.e., E[(r_i-m_i)(r_j-m_j)]). Marvell Brf. at 21-24

- Dispute
  - Should "covariance" have its ordinary meaning (Marvell) or be read out of the claim (CMU)?

circuit **32**. A noise statistics tracker circuit **34** uses the delayed samples and detector decisions to update the noise statistics, i.e., to update the noise covariance matrices. A metric computation update circuit **36** uses the updated statistics to calculate the branch metrics needed in the Viterbi-like algorithm. The algorithm does not require replacing current detectors. It simply adds two new blocks

has a feedback circuit **32** which feeds back into a Viterbi-like detector **30**. The outputs of the detector **30** are decisions and delayed signal samples, which are used by the feedback

circuit **32**. A noise statistics tracker circuit **34** uses the delayed samples and detector decisions to update the noise statistics, i.e., to update the noise covariance matrices. A metric computation update circuit **36** uses the updated statistics to calculate the branch metrics needed in the Viterbi-like algorithm. The algorithm does not require replacing current detectors. It simply adds two new blocks

in the feedback loop to adaptively estimate the branch metrics used in the Viterbi-like detector **30**.

The Viterbi-like detector **30** typically has a delay associated with it. Until the detector circuit **28** is initialized, signals of known values may be input and delayed signals are not output until the detector circuit **28** is initialized. In other types of detectors, the detector may be initialized by having the necessary values set.

The correlation-sensitive maximum likelihood sequence detector (CS-MLSD) **28** is described hereinbelow. Assume that $N>1$ channel bits (symbols), $a_1, a_2, \ldots, a_N$, are written on a magnetic medium. The symbols a_i, $i=1, \ldots, N$, are drawn from an alphabet of four symbols, $a_i \in \{+, \oplus, -, \ominus\}$. The symbols '+' and '−' denote a positive and a negative transition, respectively. The symbol '⊕' denotes a written zero (no transition) whose nearest preceding non-zero symbol is a '+' while '⊖' denotes a written zero whose nearest preceding transition is a negative one, i.e., '−'. This notation is used because a simple treatment of transitions as '1's and no transitions as '0's is blind to signal asymmetries (MR head asymmetries and base line drifts), which is inappropriate for the present problem. In FIG. **3** a sample waveform is illustrated. The signal asymmetries and base line shifts are exaggerated in FIG. **3**. FIG. **3** also shows the written symbols $a_1, \ldots, a_{18}$, as well as the samples $r_1, \ldots, r_{18}$ of

function) is factored into a product of conditional pdfs:

$$f(r_1, \ldots, r_N \mid a_1, \ldots, a_N) = \prod_{i=1}^{N} f(r_i \mid r_{i+1}, \ldots, r_N, a_1, \ldots, a_N) \tag{2}$$

To proceed and obtain more concrete results, the nature of the noise and of the intersymbol interference in magnetic recording is exploited.

Finite correlation length. The conditional pdfs in Equation (2) are assumed to be independent of future samples after some length $L \geq 0$. L is the correlation length of the noise. This independence leads to:

$$f(r_i \mid r_{i+1}, \ldots, r_N, a_1, \ldots, a_N) = f(r_i \mid r_{i+1}, \ldots, r_{i+L}, a_1, \ldots, a_N) \tag{3}$$

Finite intersymbol interference. The conditional pdf is assumed to be independent of symbols that are not in the K-neighborhood of $r_i, \ldots, r_{i+L}$. The value of $K \geq 1$ is determined by the length of the intersymbol interference (ISI). For example, for PR4, $K=2$, while for EPR4, $K=3$. $K_l \geq 0$ is defined as the length of the leading (anticausal) ISI and $K_t \geq 0$ is defined as the length of the trailing (causal) ISI, such that $K=K_t+K_l+1$. With this notation the conditional pdf in (3) can be written as:

$$f(r_i \mid r_{i+1}, \ldots, r_{i+L}, a_1, \ldots, a_N) = f(r_i \mid r_{i+1}, \ldots, r_{i+L}, a_{i-K_l}, \ldots, a_{i+L+K_t}) \tag{4}$$

Substituting (4) into (2) and applying Bayes rule, the factored form of the likelihood function (conditional pdf) is obtained:

Figure 2: Correlation between Tests #1 and #2:

This correlation value is hard to place in context, so there is a related statistic, called covariance, which measures both the degree and the direction of the relationship between two sets of data. *See* Proakis Decl. at ¶ 22; P. Olofsson, Probability, Statistics and Stochastic

covariance, which measures both the degree and the direction of the relationship between two sets of data. *See* Proakis Decl. at ¶ 22; P. Olofsson, Probability, Statistics and Stochastic

Processes, at 200-201 (2005) (Exh. 19). The covariance is zero if two sets of data are not related, positive if larger data values in one set correspond to larger values in the other set, and negative if larger data values in one set correspond to smaller data values in the other set. Olofsson at 201. The "stronger" the dependence between two sets of data, the larger the covariance. *Id.* As its name suggests, covariance is a combination of correlation and variance. Covariance is calculated by subtracting the mean for each data point to obtain deviations before the multiplication and then taking the average of the products. Pocket Dictionary of Statistics ("covariance") (Exh. 16); Ross (Exh. 17 at 119). In the test score example, the mean and deviations for Test #1 were calculated above. The mean of Test #2 is 83, and the deviations for both Test #1 and Test #2 are shown in Table 2 below. To calculate the covariance of these two

Claim Language

- "Covariance" is used independently without "noise covariance matrices"

10. A method of generating a branch weight for branches of a trellis for a Viterbi-like detector, wherein the detector is used in a system having Gaussian noise, comprising:

selecting a plurality of signal samples, wherein each sample corresponds to a different sampling time instant;

calculating a first value representing a logarithm of a quotient of a determinant of a trellis branch dependent covariance matrix of said signal samples and a determinant of a trellis branch dependent covariance matrix of a subset of said signal samples;

calculating a second value representing a quadratic of said signal samples less a plurality of target values normalized by a trellis branch dependent covariance of said signal samples;

calculating a third value representing a quadratic of a subset of said signal samples less a plurality of channel target values normalized by a trellis branch dependent covariance of said subset of signal samples;

calculating the branch weight from said first, second, and third values; and

'839 Patent at Claim 10

Claim Language

- Covariance Matrices used to calculate correlation-sensitive branch metrics

16. A method for detecting a sequence that exploits the correlation between adjacent signal samples for adaptively detecting a sequence of symbols through a communications channel having intersymbol interference, comprising the steps of:

(a) performing a Virterbi-like sequence detection on a plurality of signal samples using a plurality of correlation sensitive branch metrics;

(b) outputting a delayed decision on the transmitted symbol;

(c) outputting a delayed signal sample;

(d) adaptively updating a plurality of noise covariance matrices in response to said delayed signal samples and said delayed decisions;

(e) recalculating said plurality of correlation-sensitive branch metrics from said noise covariance matrices using subsequent signal samples; and

(f) repeating steps (a)–(e) for every new signal sample.

See '839 Patent Claims 11, 16, 19, 23; '180 Patent Claim 6

circuit 32. A noise statistics tracker circuit 34 uses the delayed samples and detector decisions to update the noise statistics, i.e., to update the noise covariance matrices. A metric computation update circuit **36** uses the updated statistics to calculate the branch metrics needed in the Viterbi-like algorithm. The algorithm does not require replacing current detectors. It simply adds two new blocks

has a feedback circuit **32** which feeds back into a Viterbi-like detector **30**. The outputs of the detector **30** are decisions and delayed signal samples, which are used by the feedback circuit **32**. A noise statistics tracker circuit **34** uses the delayed samples and detector decisions to update the noise statistics, i.e., to update the noise covariance matrices. A metric computation update circuit **36** uses the updated statistics to calculate the branch metrics needed in the Viterbi-like algorithm. The algorithm does not require replacing current detectors. It simply adds two new blocks in the feedback loop to adaptively estimate the branch metrics used in the Viterbi-like detector **30**.

The Viterbi-like detector **30** typically has a delay associated with it. Until the detector circuit **28** is initialized, signals of known values may be input and delayed signals are not output until the detector circuit **28** is initialized. In other types of detectors, the detector may be initialized by having the necessary values set.

The correlation-sensitive maximum likelihood sequence detector (CS-MLSD) **28** is described hereinbelow. Assume that N>1 channel bits (symbols), $a_1, a_2, \dots, a_N$, are written on a magnetic medium. The symbols a_i, $i=1, \dots, N$, are drawn from an alphabet of four symbols, $a_i \in \{+, \oplus, -, \ominus\}$. The symbols '+' and '−' denote a positive and a negative transition, respectively. The symbol '⊕' denotes a written zero (no transition) whose nearest preceding non-zero symbol is a '+' while '⊖' denotes a written zero whose nearest preceding transition is a negative one, i.e., '−'. This notation is used because a simple treatment of transitions as '1's and no transitions as '0's is blind to signal asymmetries (MR head asymmetries and base line drifts), which is inappropriate for the present problem. In FIG. **3** a sample waveform is illustrated. The signal asymmetries and base line shifts are exaggerated in FIG. **3**. FIG. **3** also shows the written symbols $a_1, \dots, a_{18}$, as well as the samples $r_1, \dots, r_{18}$ of

function) is factored into a product of conditional pdfs:

$$f(r_1, \dots, r_N | a_1, \dots, a_N) = \prod_{i=1}^{N} f(r_i | r_{i+1}, \dots, r_N, a_1, \dots, a_N) \tag{2}$$

To proceed and obtain more concrete results, the nature of the noise and of the intersymbol interference in magnetic recording is exploited.

Finite correlation length. The conditional pdfs in Equation (2) are assumed to be independent of future samples after some length L≧0. L is the correlation length of the noise. This independence leads to:

[illegible] (3)

Finite intersymbol interference. The conditional pdf is assumed to be independent of symbols that are not in the K-neighborhood of $r_i, \dots, r_{i+L}$. The value of K≧1 is determined by the length of the intersymbol interference (ISI). For example, for PR4, K=2, while for EPR4, K=3. $K_l \geq 0$ is defined as the length of the leading (anticausal) ISI and $K_t \geq 0$ is defined as the length of the trailing (causal) ISI, such that $K=K_l+K_t+1$. With this notation the conditional pdf in (3) can be written as:

[illegible] (4)

Substituting (4) into (2) and applying Bayes rule, the factored form of the likelihood function (conditional pdf) is obtained:

Claim Language: Independent Meaning

- "Covariance matrix" is used independently without "noise covariance matrices"

10. A method of generating a branch weight for branches of a trellis for a Viterbi-like detector, wherein the detector is used in a system having Gaussian noise, comprising:

selecting a plurality of signal samples, wherein each sample corresponds to a different sampling time instant;

calculating a first value representing a logarithm of a quotient of a determinant of a trellis branch dependent covariance matrix of said signal samples and a determinant of a trellis branch dependent covariance matrix of a subset of said signal samples;

calculating a second value representing a quadratic of said signal samples less a plurality of target values normalized by a trellis branch dependent covariance of said signal samples;

calculating a third value representing a quadratic of a subset of said signal samples less a plurality of channel target values normalized by a trellis branch dependent covariance of said subset of signal samples;

calculating the branch weight from said first, second, and third values; and

'839 Patent at Claim 10

Specification

- Covariance Matrix C_i is a "Matrix."

As stated above, the covariance matrix is given as:

$$C_i = \begin{bmatrix} \alpha_i & c_i \\ c_i^T & C_i \end{bmatrix}. \quad (14)$$

'839 Patent 7:25-29

the data dependent nature, of the circuit **52**. The weights $\underline{w}_i$ and the value $\sigma_i{}^2$ can be adapted using three methods. First, $\underline{w}_i$ and $\sigma_i{}^2$ can be obtained directly from Equations (20) and (16), respectively, once an estimate of the signal-dependent covariance matrix C_i is available. Second, $\underline{w}_i$ and $\sigma_i{}^2$ can be calculated by performing a Cholesky factorization on the inverse of the covariance matrix C_i. For example, in the $L_i D_i^{-1} L_i^T$ Cholesky factorization, $\underline{w}_i$ is the first column of the Cholesky factor L_i and $\sigma_i{}^2$ is the first element of the diagonal matrix D_i. Third, $\underline{w}_i$ and $\sigma_i{}^2$ can be computed directly from the data using a recursive least squares-type algorithm. In the first two methods, an estimate of the covariance matrix is obtained by a recursive least squares algorithm.

'839 Patent 8:10-23

Computing the branch metrics in (10) or (13) requires knowledge of the signal statistics. These statistics are the mean signal values m_i in (12) as well as the covariance matrices C_i in (13). In magnetic recording systems, these

'839 Patent 8:24-27

Noise Covariance Matrices

Claim Term	CMU's Construction	Marvell's Construction
noise covariance matrices '839 Patent Claims 11, 16, 19, 23 '180 Patent Claim 6	noise statistics used to calculate the 'correlation-sensitive branch metrics.' CMU Brf. at 27	covariance matrices of signal samples (where the signal samples include noise). Marvell Brf. at 27-32

- The Dispute
  - Does "noise covariance matrices" have its ordinary meaning (Marvell) or has it been re-defined in the patent (CMU)?

Specification

- First reference to "the noise covariance matrices"
- Implies the ordinary definition

SUMMARY OF THE INVENTION

•
•
•

Because the noise statistics are non-stationary, the noise sensitive branch metrics are adaptively computed by estimating the noise covariance matrices from the read-back data. These covariance matrices are different for each branch of the tree/trellis due to the signal dependent structure of the media noise. Because the channel characteristics in magnetic recording vary from track to track, these matrices are tracked on-the-fly, recursively using past samples and previously made detector decisions.

'839 patent 2:15-23

CMU's "i.e." Argument Fails

- CMU reads out "covariance matrices" by reading in redundant language ("correlation-sensitive branch metrics")

6. A method of detecting a sequence that exploits a correlation between adjacent signal samples for adaptively detecting a sequence of symbols through a communications channel having intersymbol interference, comprising:

(a) performing sequence detection on a plurality of signal samples using a plurality of correlation sensitive branch metrics;

(b) outputti ... ymbol;

statistics used to calculate the correlation-sensitive branch metrics

(c) outputting a delayed signal sample;

(d) adaptively updating a plurality of noise covariance matrices in response to the delayed signal samples and the delayed decisions;

(e) recalculating the plurality of correlation sensitive branch metrics from the noise ~~covariance matrices~~ using subsequent signal samples; and

(f) repeating steps (a)–(e) for every new signal sample.

See '839 Patent Claims 11, 16, 19, 23; '180 Patent Claim 6

CMU's "i.e." Argument Fails

- Re-defining "noise covariance matrices" as "noise statistics" removes "covariance" and "covariance matrices" from the claims
  - "[a] claim construction that gives meaning to all the terms of the claim is preferred over one that does not do so."

    Merck & Co., Inc. v. Teva Pharm. USA, Inc., 395 F.3d 1364, 1372 (Fed. Cir. 2005).
  - "covariance" and "covariance matrices" have well-known meanings in engineering and statistics
  - "covariance" and "covariance matrices" are used independently in other claims

    See '839 Patent Claim 10.